INVISQUE: Technology and Methodologies for Interactive Information Visualization and Analytics in Large Library Collections

B.L. William Wong, Sharmin(Tinni) Choudhury, Chris Rooney,Raymond Chen, Kai Xu

Interaction Design Center, School of Engineering and Information Sciences, Middlesex University, Hendon, London, NW4 4BT England
{w.wong, t.choudhury, c.rooney, r.chen,k.xu} @mdx.ac.uk

Abstract. When a user knows exactly what they are looking for most library systems are adequate for their needs. However, when the user's information needs are ill-defined - traditional library systems prove inadequate. This is because traditional library systems are not designed to support sense making rather for information retrieval. Visual analytics is the science of analytical reasoning facilitated by interactive visualizations and visual analytics systems can support both sense making and information retrieval. In this paper, we present INVISQUE – an approach and experimental software for interactive visual search and query. INVISQUE uses an index card metaphor to displaylibrary content, organized in a way that visually integrates attributes such citations and date published, making it easy to pick out the most recent and most cited paper. It uses design techniques such as focus+context to reveal relationships between documents, while avoiding the "what-was-I-looking-for?" problem.

Keywords: Visual Analytics, Information Visualization, User Interface, Interactive Visualization

1 Introduction

Visual analytics is the science of analytical reasoning facilitated by interactive visualizations [1]. Visual analytics combines automated analysis techniques with interactive visualizationsof large and complex data sets for supporting understanding, reasoning and decision making[2]. Navigating the large collections of most digital libraries can be a very complex task, especially when approaching suchcollections with an ill-defined information need.

In contrast, well-defined information need is knowing exactly what you want,e.g.I want a book on visual analytics titled "Illuminating the Path";whereas an ill-defined information need is only having a vague idea of what you are searching for, e.g. I want information on visual analytics but I am not sure what is out there and what will be useful to me. When the information need is well-defined, information-seeking becomes a simple information retrieval (IR) task [3]. However, when the information needs are for more complex mental activities such as learning and decision making, IR is necessary but not sufficient [3].Existing library systems are well adapted at IR but they are not adequate for the far more complex task of information exploration and sense-making that is necessary when a user's information need is ill-defined and they are learning as they are exploring the information to (1) define exactly what their information needs are, (2) before they do the actual retrieval.

The JISC-funded User Behaviour in Resource Discovery (UBiRD)[4] studyinvestigated whyexpensive electronic library resources where underutilized. The study found that many electronic resource discovery systems currently used within libraries (e.g. Emerald, ISI, etc) distract users from focusing on the content, analysis and evaluation that would help them learn and make sense of the resources discovered [4].A visual analytics capable digital library system, such as the one proposed in this paper, has the potential to allow users to focus on content, analysis and evaluation and therefore, aid them to learn and make sense of what they discovered.

INVISQUE[5], Interactive Visual Search and Query Environment, developed based on findings of the UBiRD study, is an approach and experimental software for interactive visual search and query that can function as a visual analytics library system that encourages users to focus on content, analysis and evaluation. INVISQUE uses an index card metaphor for displaying digital library content. The INVISQUE visualization, which employs index-card visualisation, attempts to assist user by supporting in-context viewing by clustering index-cards by subject and by mapping information against visually integrated dimensions by ordering the index-cards on both the X and Y axes. In Figure 1, the cards are organized according to the 1st Author of the paper on the X axis and the year published on the Y axis.

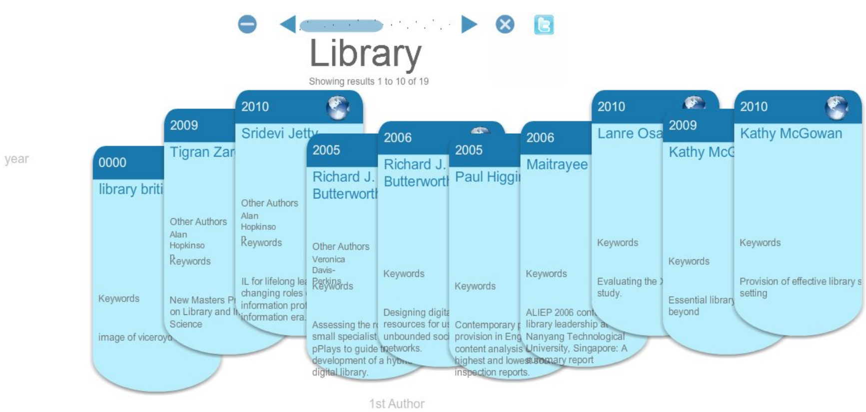

Figure 1: INVISQUE Interface

INVISQUEalso uses information design techniques such as focus+context to present detailed information while retaining the context of the search within the visual field of view. This helps users maintain their orientation within various searches, minimize their chance of getting lost in the dataand experiencing the "what-was-I-looking-for?" or 'WWILF-ing' problem, which is a problem with the current search systems [6]. INVISQUE also attempts to support higher order functions such as finding semantically meaningful relationships between data entities by revealing resources sharing the same or semantically similar relationships (Fig. 3 and 4), when prompted to do so [7].

In this paper, we first discuss the background and motivations behind INVISQUE before illustrating the functional capabilities of INVISQUE using a use case scenario. We then discuss the system structure and implementation of INVISQUE before presenting our planned future work and concluding.

2 Background & Motivation

INVISQUE was developed as a possible solution to the problems identified in the UBiRD study: poor usability, high complexity, and lack of integration in many electronic resource discovery systems, acting as a barrier to information search and retrieval[4].The studyalso found that users did not understand how to assess the quality of materials they found, which theUBiRD researchers attributed to poor information literacy[4]. However, Spink offers a different view and suggests that the user's ability to rank relevance and irrelevance of information is based on 1) information problem definition, 2) search intermediaries' perceptions that a user's question and information problem has changed during the mediated search interaction, 3) personal knowledge due to the search interaction, and 4) criteria for making relevance judgments [8], i.e. the more well-defined users' needs are, the more they are able to judge relevance. By contrast, the more ill-defined the users' needs are the more they need guidance in order to judge relevance. The study also observedthat the level of the user's domain knowledge alters their behavior [4].

Both Spink and the UBiRD study observed a process of progressive understanding from the perspective of the user where the user often startswith an ill-defined search criterion in their mind but continuously refines the criteria based on the results they obtain from IR tools[4,9,10].In addition, studies conducted by Kodagoda on the search behavior of low and high literacy users, showed that low and high literacy users demonstrated markedly different search behaviors with low literacy being at a disadvantage with conventional list form result displays [11]. Motivated by these studies, INVISQUE was developed as a potential solution to the issues identified [5].

Combining multidimensional visualisations and dynamic queries is not a new concept. Ahlberg and Shneiderman visualized search results using two dimensional scattergrams and provided sliders to filter the data [12,13]. HomeFinder used dynamic queries and sliders for user to controlvisualization of multidimensional data[14]. More recently Stasko et al., developed the JIGSAWsystem that provides multiple coordinated views of document entities emphasizing visual connections between entities across the differentdocuments[15]. INVISQUE combines the information visualization concepts mentioned above with a modern visual interface and emerging interaction technologies with the goal being to assist users the sensemaking process.

INVISQUE is heavily influenced by the Pirolli& Card modelof intelligence sensemaking, shown in Figure 2, [16]. The Pirolli& Card model came out of analysis of sensemakingactivities that took place amongst intelligence analyst [16]. However, the model can be appliedto sensemaking in other domains, such as the digital library domain.

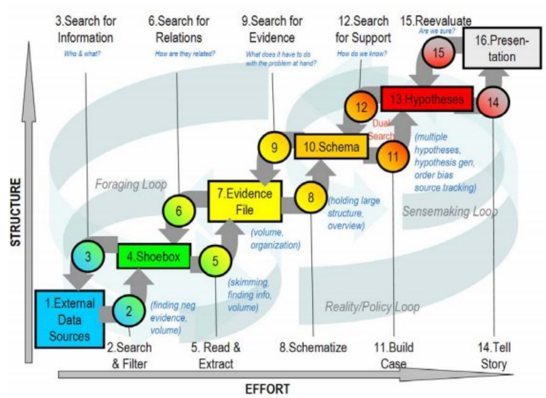

Figure 2: Pirolli& Card Model Of Sense-making[16]

The INVISQUEsystem is supported by a hybrid adaptive architecture[17]. An adaptive architecture changes its structure based on use and demand [18]. Advocates of the architecture propose adaptive system based on adaptive software architecture to be the key to achieving the goal of retaining full application plasticity throughout the software's lifecycle and that are as easy to modify on the field as they are on the drawing board [18]. We aim to take advantage of adaptive architectures to make INVISQUE more robust, scalable and changeable.

3 INVISQUE Interface – An Use Case Demonstration

The best way to highlight the properties of INVISQUE is through a use case. Our INVISQUE demonstrator has been connected to a range of different data sets such as the National Counter Terrorism Center's (NCTC) Worldwide Incident Tracking System (WITS) dataset, Google cinemas and e-gov social services datasets. For this paper, our use case describes INVISQUE connected to the Middlesex University ePrints Repository and the typical information search of an early-stage PhD.Firstly, while the PhD student will become an expert of sorts in their chosen field by the end of their PhD candidature, at the beginning of said PhD candidature – the student is a novice with their very topic of their PhD changing and evolving and thus the student's information needs are highly ill-defined.

Secondly, the PhD student would fall in the category of people who, according to Spink and the findings of UBiRD[4,8] would not necessarily know what is relevant and what is irrelevant, adding to thechallenges in ill-defined nature of the students query. Lastly, as defined by Marchionini, the PhD student is engaged in complex mental activities of learning and decision makingwhile exploring the literature [3].

Let us assume that this student is doing their PhD in interaction design under the supervision of Wong. Very likely, the first search terms they would enter into an IR tool is "wong" and "design".Figure 3 shows the two clusters of index cardsthat are produced from these two searches. By having both search results in the same visual field of view, the student immediately engages in sensemaking by explore relationships through the focus+context features of INVISQUE. As shown in Figure 3, when an index card is selected – other cards which share traits with the highlighted card come into focus, while non-related cards fade into the background. This gives the students information about relationships that they would not easily get from a list display, and assists in the sense-making process by highlighting common traits such as shared keywords, authors and publication years. In Figure 3, the common trait being highlighted across the two clusters is the author.

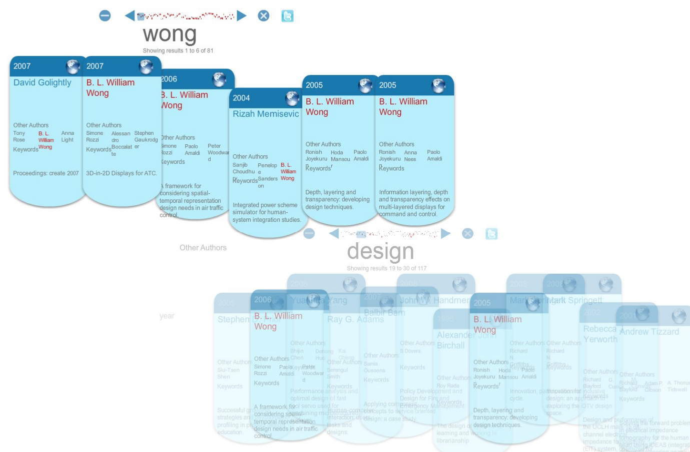

Figure 3: Display Showing Two Result Clusters With Focus+Context

Now armed with more information the student can launch a third search based on, for example, keyword discovered amongst the first set of results. Alternatively, the student can start “shoeboxing”, see Figure 2, gathering and categorizing publications they are interested. INVISQUE also allows the gathering of selected cards into a new cluster.The student can also customize the X-Y axes to specify the manner in which the index cards are arranged or zoom into the *Contextual Interval Slider*, shown in Figure 4, whichprovides the student an indication of how many other records within the dataset that share the same traits. In addition, the student also has the ability to drill down and bring-up addition information about a selected index card – for example, the PDF of a journal article.

Figure 4: Records Of Interest Highlighted In The Contextual Interval Slider

INVISQUE embodies the concept of an infinite canvas which allows for a potentially limitless amount of search clusters to be displayed together. The infinite canvas concept has the potential of being adapted and deployed into collaborative environments, enabling multiple users to search simultaneously.

4 INVISQUESystem Architecture

INVISQUE is very much a working prototype with new features and functionality being added to the system almost on a daily basis. It is also not just a digital libraries visual analytics system. We have put INVISQUE to other uses, such as its ability to act as an information kiosk for entertainment information [19]. The entertainment information kiosk version of INVISQUE differs mainly in the interface, with more emphasis being placed on multi-touch based user interactions as opposed to the point and click mouse-based interaction emphasized in the digital library version of INVISQUE. However, while the interface differs significantly between deployments of INVISQUE – the overall system architecture is designed to remain consistent. Indeed, one of the reasons we opted for an adaptive architecture is so that INVISQUE can change depending on the use to which it is placed.

In Section 2, we called the INVISQUE architecture a hybrid adaptive architecture. The hybridization is with a 3-tier client-server architecture which forms the primary structure of the INVISQUE architecture. In the "middle-tier" of the n-tier architecture, we have imbedded a rule-based *Architecture Controller* that orchestrates functional components based on data-type, data-load and functional prompts from the interface. The architecture is illustrated in Figure 5.

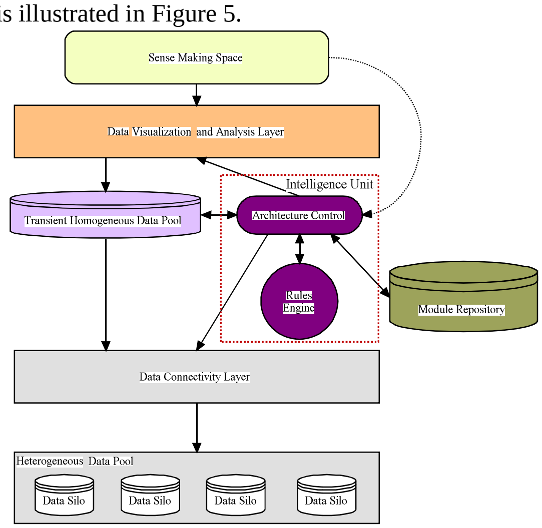

Figure 4: INVISQUE Architecture

To explain the architecture, at the data-tier level we have a number of silos containing heterogeneous data. The UBiRD study revealed that users prefer integrated systems [4] and INVISQUE does aim to provide an integrated view of multiple data sources. Therefore, the first layer of functionality in the middle-tier is the *Data Retrieval Component Orchestration Layer*. The exact functionalities are deliberately abstracted and will be determined by the data silos the system has to access.

For example, the Middlesex University ePrints Repository to which INVISQUE is currently connected to is a MySQL database. However, we are also working on a project that connects INVISQUE to data that is in XML format. INVISQUE will work with both the MySQL database and the XML database in tandem and display a merged result set when a keyword search is executed. This is where the *Transient Homogenous Result Pool* comes in. The idea is that results from underlying data silos will be gathered in the pool and made available to the upper tier functions; so that the upper tier do not have to consider data structure. Additionally, this separation also means that the *Architecture Controller* has greater control over adaptation.

Once the results are in a collected format, the INVISQUE interface is free to execute operations on the result pool. Again, the*Data Visualization and Analysis Component Orchestration Layer* is kept deliberately abstracted because the idea is to add and remove functionality quickly and easily as required, thus scaling and changing the system dynamically. This is because INVISQUE is only at a fraction of what it is capable of. So that we can easily add and remove functionality and adapt the same system for multiple uses, the functional layer has to be as abstract as the data retrieval layer with the nucleus of the architecture, the *Architecture Controller*, pulling from a pool of components the required components and orchestrating them on the relevant layer based on predefined rules that are triggered by both interactions happening in the user interface and changes in the data-tier.

5 Evaluation &Summary

At the time of writing, as part of her PhD, Kodagoda had evaluated INVISQUE on high and low literacy users in the context of finding on-line social service information [6].Kodagoda tested the hypothesis that the context layering offered by INVISQUE would reduce premature search abandonment when compared with traditional hierarchical website layout. The results to dateare in favor of INVISQUE[6].In the near future we will be conductingscalability tests and re-design once we complete the integration of a 2 terabyte journal archive provided to us by a major journal publisher, as mentioned in Section 4.

INVISQUE aims to present the design for the next generation of information search and retrieval systems that would support semantic analysis and access to massively large data sets. The design we have developed shows how we can visualize and present search results and we can facilitate interaction with the visualization.

We have moved away from the conventional list-style arrangement, and instead represent information by the use of index cards in a 2-dimentional space. The open canvas allows users to perform multiple searches while keeping the context of the complete search space.

We believe that INVISQUE's visibly novel user interface, backed-up with its hybrid adaptive architecture can overcome the shortcoming of existing library systems.

Acknowledgments.INVISQUE was originally funded by the JISC Rapid Innovation program, Grant Ref. Num. IEDEVC19/RI. We would also like thank Middlesex University for its continued supportas well as the Interaction Design Center team who have had an input into INVISQUE.

References

[1] J.J. Thomas and K.A. Cook, *Illuminating the Path: The research and development agenda for visual analytics*, IEEE Computer Society, 2005.

[2] D. Kei, G. Andrienko, J.-D. Fekete, C. Gorg, J. Kohlhammer, and G. Melancon, “Visual Analytics: Definition, Process, and Challenge,” 2008.

[3] G. Marchionini and R.W. White, “Information-seeking support systems,” *IEEE Computer*, vol. 42, 2009, pp. 30-32.

[4] W. Wong, H. Stelmaszewska, N. Bhimani, S. Barn, and B. Barn, *User Behaviour in Resource Discovery: Final Report*, 2009.

[5] Nawaz Khan, *INVISQUE, INteractiveVIsual Search and QUery Environment*, 2010.

[6] N. Kodagoda, “PhD Thesis of NeeshaKodagoda,” Middlesex University, 2011.

[7] H. Stelmaszewska, B.L.W. Wong, S. Attfield, and R. Chen, “Electronic resource discovery systems: from user behaviour to design,” *Proceedings of the 6th Nordic Conference on HumanComputer Interaction Extending Boundaries*, ACM, 2010, pp. 483-492.

[8] A.H. Spink, H. Greisdorf, and J. Bateman, “From highly relevant to not relevant: examining different regions of relevance,” *Information Processing and Management*, vol. 34, 1998, pp. 599-622.

[9] A. Spink, D. Wolfram, M.B.J. Jansen, and T. Saracevic, “Searching the web: The public and their queries,” *Journal of the American Society for Information Science and Technology*, vol. 52, 2001, pp. 226-234.

[10] A. Spink and T. Saracevic, “Human-computer interaction in information retrieval: nature and manifestations of feedback,” *Interacting with Computers*, vol. 10, 1998, pp. 249-267.

[11] N. Kodagoda, W.B.L. Wong, and N. Khan, “Information seeking behaviour model as a theoretical lens: high and low literate users behaviour process analysed,” *ECCE*, 2010, pp. 117-124.

[12] C. Ahlberg and B. Shneiderman, “Visual Information Seeking : Tight Coupling of Dynamic Query Filters with Starfield Displays,” *ACM CHI Conference on Human Factors in Computing Systems*, Boston, USA: 1994, pp. 313-321.

[13] B. Shneiderman and C. Ahlberg, “AlphaSlider: A compact and rapid selector,” *ACM CHI Conference on Human Factors in Computing Systems*, 1994.

[14] C. Williamson and B. Shneiderman, “The Dynamic HomeFinder: Evaluating dynamic queries in a realestate information exploration system,” *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval*, 1992.

[15] J. Stasko, C. Görg, and R. Spence, “Jigsaw: supporting investigative analysis through interactive visualization,” *Information Visualization*, vol. 7, 2008, pp. 118-132.

[16] P. Pirolli and S. Card, “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis,” *Proceedings of the 2005 International Conference on Intelligence Analysis*, 2005.

[17] S.T. Choudhury, “Loculus: An ontology-based information management framework for the Motion Picture Industry,” Queensland University of Technology, 2010.

[18] P. Oreizy, M.M. Gorlick, R.N. Taylor, D. Heimbigner, G. Johnson, N. Medvidovic, A. Quilici, D.S. Rosenblum, and A.L. Wolf, “S E L F - A D A P T I V E An Architecture-Based Approach to Self-Adaptive Software,” *IEEE Intelligent Systems*, 1999, pp. 54-62.

[19] W. Wong, C. Rooney, R. Chen, K. Xu, and N. Kodagoda, “INVISQUE : Intuitive Information Exploration through Interactive Visualization,” *ACM CHI Conference on Human Factors in Computing Systems*, 2011.